

Reading Skill and Test Design Principles: Designing a Diagnostic Reading Comprehension Test for Kurdish EFL learners in the University Level

Atta Mostafa Hamamrad

College of Basic Education, University of Halabja, Kurdistan Region - Iraq

Abstract

Tests have always been considered as an indispensable part of education. Test, with its different types, has been the criterion to evaluate and decide on learners' outcome and performance in the learned subjects. This paper sheds light on the factors that are essential to be taken into consideration during designing any written test. It also presents a framework for a reading comprehension test that can be designed and administered for EFL Kurdish learners in the university level.

Keywords: reading comprehension, diagnostic test, test principles

1. Introduction

Language testing history is in part, the history of efforts and attempts made to fill the gap between test and real-life usage of language. Tests are designed to evaluate learners' abilities in different language aspects and provide testers with information concerning learners' language proficiency development and level of their understanding of the taught material. Testing is a way of collecting evidence about learners' progress and response to the teaching learning process.

2. Reading comprehension and designed test specifications

This section renders an account regarding reading comprehension and related theories in addition to explaining the purpose of the test, the test construct, and some samples taken from the designed test.

2.1 Reading

Alderson (2000 p.28) defines reading as "...an enjoyable, intense, private activity, from which much pleasure can be derived, and in which one can become totally absorbed". Reading is probably the skill which is tested more frequently among the four language skills. It seems to be the easiest skill to examine. Nevertheless, it has its own difficulties.

When test takers do speaking and writing during a test, we can hear and see but when they listen or read, there will often be nothing to observe. Therefore, task of the language tester is to set the reading tests items or tasks in which result in behaviour that manifests their successful completion. It is beneficial and helpful for language tester to be familiar with theories about reading. Reading skill like other language skills has been a matter of research and scrutiny by experts in the field and different theories concerning this skill have been proposed which the most important ones are as follow:

- **Mental model theory**

This theory implies that while readers read a text, a mind film will be generated in their head based on the content of the text that they are reading.

This process often happens when the reader reads a fiction text. He/she concentrates on the main characters and creates a mental model of the circumstances in which the character finds herself/himself. As the situation changes, the mental model is reconstructed to update and mirror the new circumstances but the important items related to the main character (s) is (are) kept in the foreground (Gunning 1996).

- **Schemata theory**

According to Gunning (1996) schemata is the organised existing knowledge that a person has about people, events, things and places. Schemata theory proposes an interaction between the text and reader's existing knowledge which leads to understanding the text (Kitao & Kitao 1995). This schemata or existing knowledge may be wide and extensive like knowledge of natural disasters or can be narrow and limited like knowledge of volcano. While students read materials (texts), they can associate the new information of the texts to their already existing knowledge which is compartmentalized and stored in their mind. Therefore, they can comprehend the texts and meanwhile, they can add this information to their stored knowledge for future usage. This theory suggests that students' degree of comprehension may vary according to the extension of their information compartments.

- **Proposition theory**

As this theory proposes, reader constructs a macrostructure or a main idea as he/she is processing the text. These macrostructures and ideas are arranged in hierarchy and things are given priority to be memorised according to their importance and most significant things are given the highest priority (Gunning 1996). Based on these theories, it can be said the a successful comprehensible reading, especially while doing a reading test, requires some micro and macro skills on the side of reader or test taker. Accordingly, a good designed reading test should involve these skills of candidates or the testees.

Hughes (2003 pp.116-117) furthermore, identifies two types of skills which are essential for learners to consider during doing any reading comprehension test:

A) Macro-skills such as;

1. *Scanning text to locate specific information*
2. *Skimming to get the gist*
3. *Identifying stages of an argument*
4. *Identifying examples given in support of the argument.*

B) Micro-skills such as;

1. *Using context to guess the meaning of unfamiliar words*
2. *Identifying references of pronouns*
3. *Identifying relation between different parts of the text by recognizing indicators in discourse to understand the introduction, development and conclusion idea.*

Thus, the test can address students' cognitive capability of constructing meaning based on what is entailed in the text. It also assists students to develop their reading skills and reading strategies as well as reading fluency. Furthermore, test honours students' literacy ability and context to make affective and cognitive leaps based on a synthesis of old and new information (Wiggins 1990).

The choice of text on the other hand, is a vital and integral part of designing a reading test. Teachers should use the kinds of texts as input for their reading test which are in line with teaching syllabus and program purpose, suit their interest and language proficiency level and familiar to students. However sometimes the teacher can use a kind of text along with the familiar text which is new and a little higher than students' predictive language proficiency level so that he can differentiate between students' proficiency level and know to which extent students are able to deal with new, less familiar texts and tasks.

2.2 The purpose of the designed test

The test intends to identify the weaknesses and strengths of the students with regard to reading for comprehension subject. Also, to assess their ability to grasp the gist, understand details, find specific information, guess the meaning of lexical items in context and to draw inferences from a text. This test is carried out at the end of the first semester in order for the teacher to have a clearer understanding of students' reading abilities and take suitable curative steps to treat their weaknesses for the second semester.

2.3 Test construct

It is a diagnostic reading test and aims at measuring abilities of students in terms of reading strategies by providing them with different text materials. The intended reading strategies to be used for answering this test are 'skimming and scanning' (trying to find specific details e.g. Figures, names, dates, and so on) and 'careful reading' (understanding the main ideas stated in the text, and also the ability to make propositional inferences by deducing implicit information from explicit information of the text as well predicting the meaning of words based on context).

2.4 Samples of the designed test

The test design encompasses a number of different input texts which tackle different topics. The first text concerns politics. The second is a dialogue and the last text is an informative report. The following are some examples of the texts and the questions. The texts have been adopted from the google website.

TEXT.1 / Target reading skills: Skimming & Scanning for specific information

Read the passage below and then answer the questions by circling the correct choice letter. Write your answers on question sheet.

United States Presidential Elections

This year Americans elect on a new president on November 6th. An important event that happens once every four years. Currently, the president is always elected from one of the two main parties in the United States: the Republicans and the Democrats. There are other presidential candidates. However, it is unlikely that any of these "third party" candidates will win. It certainly has not happened in the last one hundred years.

In order to become the presidential nominee of a party, the candidate must win the primary election. Primary elections are held throughout each state in the United States in the first half of any election year. Then, the delegates attend their party convention in order to nominate their chosen candidate. Usually, as in this election, it is clear who will be the nominee. However, in the past parties have been divided and choosing a nominee has been a difficult process.

Once the nominees have been selected, they campaign throughout the country. A number of debates are usually held in order to better understand the candidates' points of view. These points of view often reflect their party's platform. A party platform is best described as the general beliefs and policies a party holds. Candidates cross the country by plane, bus, train or by car giving speeches. These speeches are often called 'stump speeches'. In the 19th century, candidates would stand on tree stumps to deliver their speeches. These stump speeches repeat the candidates' basic views and aspirations for the country. They are repeated many hundreds of times by each candidate.

Q1: How often does the United States hold presidential elections?

- A) every two years
- B) every four years
- C) every six years

Q2. Indicate whether the statement is true, false, or not given by writing the letter T (True), F (False), or NG (Not Given).

The delegations attend their party convention to nominate their chosen candidate for the primary election.

Q3. Fill in the gaps with appropriate words from the passage.

Candidates speeches are known as and they repeat the candidates' and for the country.

TEXT.2 / Target reading skill: Understanding lexis by predicting the meaning of words from the context.

Based on the information provided in the dialogue, choose the correct meaning of the underlined word by encircling the correct choice letter on your question paper.

Tom: I love your glasses!

Jim: Thanks. They are brand new.

Tom: You look great. They are very flattering.

1) Tom tells Jim the glasses are flattering. He means that the glasses...

- A. are shiny like a mirror.
- B. make Jim look good.
- C. are a good brand.
- D. Look new

TEXT.3 / Target reading skill: Making propositional inferences

Read the passage below and then choose the best answer to the question by encircling the correct choice letter on your question sheet. Answer the question based on what is stated or implied in the passage.

Naval architects never claim that a ship is unsinkable, but the sinking of the passenger-and-car ferry Estonia in the Baltic surely should have never happened. It was well designed and carefully maintained. It carried the proper number of lifeboats. It had been thoroughly inspected the day before fatal voyage. Yet, the hours later, the Estonia rolled over and sank in a cold, Stormy night. It went down so quickly that most of those on board, caught in their dark, flooding cabins, had no chance to save themselves; of those who managed to scramble overboard, only 139 survived. The rest died of hypothermia before the rescuers could pluck them from the cold sea. The final death toll amounted to 912 souls. However, there were an unpleasant number of questions about why the Estonia sank and why so many of survivors were men in the prime of life, while most of the dead were women, children and elderly.

Q1. One can understand from the reading that....

- A) The lifesaving equipment did not work well and lifeboats could not be lowered.
- B) Design faults and incompetent crew contributed to the sinking of the Estonia ferry.
- C) Most victims were trapped inside the boat as they were in their cabin.
- D) Naval architects claimed that the Estonia was unsinkable.

3. Rationale for the designed test

According to Bachman (1990 p.33) "language tests are indirect indicators of the underlying traits in which we are interested". Dochy (1992 p.26) states that test is "an often formalised (collection of) task(s), designed to determine a test taker's ability, knowledge or intelligence". Hughes (2003) considers test as a tool which helps us acquire information about a particular skill or skills of test takers. It also acts as a decision-making procedure or a procedure for problem-solving. Brown and Abeywickrama (2010) propose that tests help in ranking learners, judging if learners are capable of coping with certain language forms and deciding whether learning goals have

been achieved. Furthermore, tests provide valuable information relating to predicting learners' further progresses in the future as well as indicating the effectiveness of teacher's methodology in the classroom. He also regards test as a medium to control and motivate learners. Language testing at any level is a complicated task which should be based on theory as well as practice. Here the focus is on practical aspect of language testing and guiding principles which indicate a good test. These principles are:

3.1 Validity

Validity is "the extent to which the information you collect actually reflects the characteristic or attribute you want to know about" (Genesee & Upshur 1996 p.62). According to Hughes (2003 p.22) "a test is said to be valid if it measures accurately what it is intended to measure". Bachman on the other hand, defines validity as "the degree to which the evidence supports the inferences that are made from the scores" (1990 p.237). There are several categories of validity which are:

3.1.1 Content validity

It is concerned with the degree to which the components of a test relate to a real-life situation they attempting to replicate (Bachman 1990). It indicates that the test contains a representative sample of the language structures or skills that exist in the course content or program and are meant to be tested. This depends on the purpose of the test and the targeted skill (s) which is clarified in the test specifications. Validity content of a test which covers a representative sample of various language skills is crucial as it provides a good washback effect about effectiveness and suitability of syllabus and teaching methodology.

3.1.2 Construct validity

It is concerned with the level of accuracy a construct within a test is believed to measure (Brown 1994). Construct validity refers to the "fitness" between any underlying ability or trait which is hypothesized in a language theory and the test item. It is concerned with whether a particular item or testing technique exactly tests the ability which is supposed to be tested. For instance, reading theory proposes that several different construct (ability, trait or a psychological concept)s are involved in reading such as skimming and scanning and each one is different from the other. Therefore, any reading test item is considered to have construct validity if it indeed measures that specific construct. Construct validity is highly significant in judging the quality of standardized tests however, it is less needed in classroom based assessment as it is more direct and based on collecting several types of information (Genesee & Upshur 1996). It is advisable for test constructors to investigate what is known about testing particular constructs before writing items.

3.1.3 Face validity

It implies that the appearance of the test gives the impression as if it measures what is supposed to. Also it is viewed as "the extent to which a test appears on the surface to accurately assess what is meant to assess" (Hughes 2003 p.27). A reading test for instance, must clearly involve reading. Lacking this feature, a test would be refused by education authorities, teachers and test takers. Face validity is an important factor as a professional looking test has more credibility with administrators and students than a sloppy one.

Efforts have been exerted to consider different aspects of validity in the proposed designed test. Regarding content validity, samples which have been chosen to test the skills and language structures that exist in the university teaching course and syllabus. Students are familiar with this type of reading comprehension test and they have been doing similar tests from secondary school. Concerning validity, each part of the designed test is set to examine a particular ability or construct of the students. For instance, the first part of the test aims at measuring students' ability to skim and scan a text to find specific details about a figure (football player), dates and his achievements in football. The second and the third parts, are concerned with the detailed reading for comprehension and test students' ability to grasp the main ideas mentioned in the text and the ability to infer the results based on their understanding of the text. Face validity has also been regarded as the given test merely involves students reading skills and questions are set in a way that do not require much writing and only encircling the correct choice letter on the question sheet is sufficient.

3.2 Reliability

It refers to consistency, generalizability and stability of a test. It is also defined as "freedom of nonsystematic fluctuation" (Genesee & Upshur 1996 p.57). It conveys that a test however, given in different settings, times and in different forms, will produce similar results. For example, if we administer a test to a group of students in a setting and a time then we give the same test to the same group at a different time and setting, it should not make any, or at least not much, difference in their results. Likewise, if we interchangeably give two different forms (versions) of a test to students to answer, it should not make any difference for the students which form (version) of the test they are answering or dealing with. They should obtain approximately the same result with either form of the test.

There are several factors which influence reliability of a test such as test takers' age, gender, personal characteristics as well as test environment, test formats, content of questions and the amount of time given to

take the test (Bachman 1990). A test including more items is considered more reliable because in this case, the teachers have more samples of students' language ability. Also testing research has shown that longer test generate more reliable results than brief quizzes (Bachman 1990). Moreover, classroom setting like noise, acoustics, seating arrangements and lighting on the one hand, and anxiety, fatigue and learning style on the other hand, can affect the reliability of a test.

While developing and using language tests, it is crucially important to identify the potential threats that endanger test reliability meanwhile, try to minimise their effect. Henning (1987) reports them as follow:

- **Learner fluctuations;** a number of changes might occur with learners that may change learner's true score from test to test. Changes like forgetting or additional learning and influences like sickness, tiredness, emotional problems and practice effect (familiarity with the content of the test that he or she has taken it several times) can result in learner's score to deviate from the score that reflects her or his actual ability.
- **Test administration fluctuations;** inconsistency in testing conditions and administrative procedures will result in reduction in test reliability. This problem can commonly be seen in institutions where different students are tested in different locations on different days.
- **Scoring fluctuations;** subjectivity in test scoring is a problematic case as it involves scorer's opinion and personal judgment and this may lead to unfairness in the test results.

Reliability is considered an indispensable quality of test scores because if the scores are not relatively consistent, they cannot give a true information about learners' abilities which are being measured. "Reliability and validity are closely interwoven. While reliability concentrates on the empirical facets of the testing process, validity stresses on the theoretical facets and intertwines these concepts with the empirical ones." (Davies 1990 p.169).

To achieve a high reliability, items of designed test are mostly set in multiple choice style and a mark is allocated for each correct answer. This impedes any sort of subjectivity of the scorer. Also test parts are well laid out and clear instruction is catered for each part. Test is conducted in the classroom (the number of students would probably be around 30-40 students). Therefore, for a better administration and monitoring, an invigilator might be called to assist in the process. The approximate time for doing this test is calculated beforehand and students are given sufficient time so as to reduce their anxiety about shortage of time. Moreover, an effort will be made to create a calm atmosphere in the classroom.

3.3 Practicality

It is another important consideration when choosing or designing tests. Administrators and teachers need to consider the relation of tests to practical issues. According to Genesee & Upshur (1996 p.57) "cost, administrative time, compilation time, administrator qualifications and acceptability" are main features of a practical test. A practical test means to be within the means of financial limitations, appropriate time constraints, easy to administer and score, and easy to interpret the results. Moreover, a good practical test necessitates availability of trained and qualified graders and administrators.

The designed test is practical as its designing, printing is quite affordable, and students only need a pencil and pen to answer the questions. It can be administered by the teacher and accomplished by students within the given time limit. Different language levels have been considered and various items, short and long, have been included to involve all students with different language levels and abilities. Items are mostly in multiple-choice format and this in turn, makes the marking process easy.

3.4 Authenticity

Authenticity is a question of judgment and knowing how well a language test task corresponds to the features of an actual "real-life" language. It is the "correspondence between language test characteristics and features of target language use" (Bachman & Palmer 1996 p.23). Designing types of tasks and test items which mirror real life situations and contexts, is a very good way to motivate learners (students) to perform their language abilities. These kinds of tasks and tests encourage students to authentically utilise the target language by engrossing them in real-life contexts and situations. Authenticity is achieved by creating a match between test tasks and target language use tasks as well as involving test takers in test tasks (Bachman 1990). Brown and Abeywickrama (2010) set some guidelines for an authentic test. For them, a test is authentic if it:

1. *Contains natural language.*
2. *Makes use of meaningful and relevant topics.*
3. *Uses good contextualization of items.*
4. *Organizes items through a theme or story line.*
5. *Uses tasks that replicate real-world tasks.*

The materials being utilised for this test reflect the natural usage of language which is heard and spoken in everyday life on the streets, or on the TV, or can be read in an article on the internet, newspapers or studied in educational institutions. These texts are written or produced by real native English speakers for the real audience

in order to impart some sorts of real messages. These printed, functional writing texts contain linguistic elements that students can make use of outside their class. They provide students with information and vocabulary which they may need when they talk or discuss the related topics in the real life or they express their opinions.

3.5 Washback

Washback is referred to as the impact which tests leave on the teaching learning process. It is a term used to describe how knowledge of a test's characteristics influence the nature of any learning preceding the test (Hughes 2003). Depending on its form, washback can be positive or negative. Sometimes, Students and teachers tend to think of the negative effects of testing and regard the teaching curriculum as a 'test-driven' curriculum or a curriculum which provide them with only 'what they need to know for the test'. In contrast, positive washback (also known as guided washback) benefits administrators, teachers and students and as it presumes that testing and curriculum design are both based on clear course outcomes that are known to students and teachers/testers. If students realise that tests are markers of their progress towards attaining these outcomes, they will have a sense of accomplishment.

The washback which is received from test results will provide the teacher with valuable information about appropriateness and effectiveness of adopted pedagogy and students' strengths and weaknesses as well as helping him/her formulate suitable strategies to boost their performance in reading.

4. Conclusion

Testing is the process of evaluating learners' ability, intelligence, as well as indicating their weaknesses and strengths in a learned material. There are certain criteria for designing a test that teachers and administrators should consider in their test design so that the test will be more authentic and truly measure the ability or the aspect which is intended to be measured.

Tests should consider different language levels and individual differences of students by including various topics and materials as well containing long, short items and questions that are set for different levels of intelligence.

While designing a reading test, it is advisable that testers choose the kinds of materials and texts which are authentic and mirror the daily life and real-life situations. These materials will encourage students for a natural use of the target language.

References

- Alderson, J. C. (2000). *Assessing Reading*. Cambridge: Cambridge University Press.
- Bachman, L. F. (1990). *Fundamental considerations in language testing*. Oxford: Oxford University Press.
- Bachman, L. F. & Palmer, A. S. (1996). *Language testing in practice*. Oxford, UK: Oxford University Press.
- Brown, H. D. (1994). *Principles of Language Learning and Teaching*. (3rd ed.). Englewood Cliffs, NJ: Prentice Hall Regents.
- Brown, H. D. & Abeywickrama, P. (2010). *Language assessment, principles and classroom practice*. (2nd ed.). White Plains, NY: Pearson Education, Inc.
- Dochy, F. J. R. C. (1992). *Assessment of prior knowledge as a determinant of future learning: the use of knowledge state tests and knowledge profiles*. Utrecht/London: Lemma B. V. /Jessica Kingsley Publishers.
- Genesee, F. & J. Upshur. (1996). *Classroom-based evaluation in second language education*. Cambridge: Cambridge University Press.
- Gunning, T.G. (1996). *Creating reading instruction for all children*. Chapter 6, 192-236.
- Henning, G. (1987). *A guide to language testing: development, evaluation, research*. Cambridge, MA: Newbury House.
- Hughes, A. (2003). *Testing for Language Teachers*. Cambridge: Cambridge University Press.
- Kitao, K., & Kitao, S. K. (1995). College reading textbooks do not meet needs. *The Daily Yomiuri*, p. 7.
- Wiggins, G. (1990). The case for authentic assessment. *Practical Assessment, Research & Evaluation*, 2(2). Retrieved from <http://PAREonline.net/getvn.asp?v=2&n=2>